

The invention claimed is:

1 1. A method for generating photorealistic
2 talking heads, comprising the steps of:
3 receiving an input stimulus;
4 reading data from a first library comprising
5 images of phoneme sequences which correspond to the
6 input stimulus;
7 reading, based on the data read from the
8 first library, corresponding data from a second library
9 comprising images of a talking subject; and
10 generating, using the data read from the
11 second library, an animated sequence of a talking head
12 tracking the input stimulus.

1 2. The method of claim 1, further
2 comprising the steps of:
3 reading acoustic data from the second library
4 associated with the corresponding data read from the
5 second library;
6 converting the acoustic data into sound; and
7 outputting the sound in synchrony with the
8 animated sequence of the talking head.

1 3. The method of claim 1, wherein the data
2 read from the first library comprises parameters
3 describing mouth shapes.

4 4. The method of claim 1, wherein the data
5 read from the first library comprises one or more
6 equations characterizing mouth shapes.

7 5. The method of claim 2, wherein the data
8 read from the first library comprises one or more
9 equations characterizing mouth shapes.

1 6. The method of claim 2, wherein said
2 converting step is performed using a data-to-voice
3 converter.

1 7. The method of claim 2, wherein the data
2 read from the first library comprises segments of
3 sampled images of a talking subject.

1 8. The method of claim 2, wherein the data
2 read from the second library comprises mouth parameters
3 characterizing degree of lip opening.

1 9. The method of claim 2, wherein said
2 receiving, said generating, said converting, and all
3 said reading steps are performed on a personal
4 computer.

1 10. The method of claim 2, wherein said
2 first and second libraries reside in a memory device on
3 a computer.

1 11. The method of claim 7, wherein said
2 first library comprises an animation library, and
3 wherein said second library comprises a coarticulation
4 library.

1 12. The method of claim 7, wherein said
2 generating step is performed by overlaying the segments
3 onto a common interface to create frames comprising the
4 animated sequence.

~

1 13. A method for generating a photo-
2 realistic talking head for a text-to-speech synthesis
3 application, comprising the steps of:
4 sampling images of a subject;

5 extracting a plurality of parameters from
6 each image sample;
7 storing the image sample parameters into an
8 animation library;
9 sampling multiphone images of the subject;
10 sampling sounds associated with the
11 multiphone images;
12 extracting a plurality of parameters from
13 each multiphone image sample;
14 storing the multiphone image parameters and
15 associated sound samples into a coarticulation library;
16 reading, based on an input stimulus
17 comprising one or more phoneme sequences, parameters
18 from the coarticulation library corresponding to each
19 phoneme sequence;
20 generating, using parameters from the
21 animation library corresponding to the read parameters,
22 a sequence of animated frames, the sequence tracking
23 the input stimulus.

24 14. The method of claim 13, wherein the
25 plurality of parameters extracted from each multiphone
26 image sample comprises data describing mouth shapes.

27 15. The method of claim 13, wherein the
28 plurality of parameters extracted from each multiphone
29 image samples comprises one or more rules
30 characterizing mouth shapes.

1 16. The method of claim 13, further
2 comprising the step of:
3 timestamping the multiphone image samples and
4 sound samples.

1 17. The method of claim 13, wherein the
2 sound samples comprise samples converted from sound
3 into data by a speech recognizer.

1 18. The method of claim 13, wherein the
2 sound samples comprise samples converted from sound
3 into data by a speech recognizer.

1 19. The method of claim 16, wherein the
2 sound samples further comprise a phoneme transcript.

1 20. The method of claim 13, wherein said
2 step of sampling images of the subject is performed by
3 a video camera.

1 21. The method of claim 16, wherein said
2 step of sampling images of the subject is performed by
3 a video camera.

1 22. The method of claim 13, wherein at least
2 one of the sampled multiphone images comprises a
3 diphone image.

1 23. The method of claim 19, wherein at least
2 one of the sampled multiphone images comprises a
3 diphone image.

1 24. The method of claim 13, wherein said
2 method is performed on a personal computer.

1 25. The method of claim 21, wherein said
2 method is performed on a personal computer.

1 26. A processor-based method for generating
2 a photo-realistic talking head for a text-to-speech
3 synthesis application, comprising the steps of:
4 sampling images of a subject;
5 decomposing the subject images into a
6 hierarchy of segments;
7 writing for each segment a set of parameters
8 into memory, the segment parameter sets characterizing
9 each segment;
10 sampling a plurality of phoneme sequences;
11 writing for each phoneme sequence a set of
12 parameters into memory, the phoneme sequence parameter
13 sets characterizing each phoneme sequence;
14 reading from memory, based upon an input
15 stimulus, specific phoneme sequence parameter sets
16 corresponding to the stimulus;
17 reading from memory, based upon the specific
18 phoneme sequence parameter sets, corresponding specific
19 segment parameter sets; and
20 generating a concatenated sequence of
21 animated frames using the corresponding specific
22 segment parameter sets.

1 27. The method of claim 26, wherein said
2 generating step is performed by overlaying onto a
3 common interface, for each animated frame, a plurality
4 of segments corresponding to the specific segment
5 parameter sets.

1 28. The method of claim 26, wherein said
2 generating step comprises outputting the concatenated
3 sequence to a screen.

1 29. The method of claim 27, wherein said
2 generating step further comprises outputting the
3 concatenated sequence to a screen.

1 30. The method of claim 27, wherein the
2 segments comprise facial parts.

1 31. A method for generating a photo-
2 realistic talking head for a text-to-speech synthesis
3 application, comprising the steps of:
4 sampling images of a talking head;
5 extracting a plurality of parameters from
6 each image sample;
7 writing the image sample parameters into an
8 animation library;
9 sampling multiphone images of the subject;
10 sampling sounds associated with the
11 multiphone images;
12 converting the sound samples into digital
13 acoustic parameters;
14 extracting a plurality of parameters from
15 each multiphone image sample;
16 storing the multiphone image parameters and
17 associated acoustic parameters into a coarticulation
18 library;
19 reading, based on an input stimulus
20 comprising one or more phoneme sequences, parameters
21 from the coarticulation library associated with each
22 phoneme sequence;
23 generating, using parameters from the
24 animation library, a sequence of animated frames
25 corresponding to the read parameters and a sequence of
26 associated sounds in synchrony with the animated frames
27 sequence, the sequence of animated frames tracking
28 the input stimulus.

Cosatto -20-1

- 1 32. The method of claim 31, wherein said
- 2 converting step is performed by a speech recognizer.